**6П**

**УДК 004.382**

***С.А. Поливцев***
Институт проблем искусственного интеллекта МОН и НАН Украины, г. Донецк
psa@iai.donetsk.ua

# Многопроцессорная система реализации бинаурального слуха с массивом микрофонов

В статье предлагается вариант практической реализации системы бинаурального слуха, присущего человеку и многим представителям животного мира. Свойства бинаурального слуха реализуются путем операций над образами акустических сигналов, получаемых с массива микрофонов. Реализация предполагает связную, синхронизированную работу 3 микропроцессоров, выполняющих чтение сигналов с 4 микрофонов через два 16-разрядных АЦП с частотой дискретизации 100 кГц по каждому каналу с синхронизацией между каналами с точностью 5 мкС. В реализацию включена FLASH карта емкостью 2 GB и микропроцессор, реализующий интерфейс USB с персональной ЭВМ. Реализация предусматривает как автономную работу системы, так и совместно с персональной ЭВМ. Система имеет свойства, не присущие слуху человека, – нулевое время адаптации при переходе от сильного сигнала к слабому и наоборот, локализацию очень коротких сигналов.

## Введение

Под бинауральным слухом понимается система восприятия и анализа звуковых сигналов окружающей среды, подобная той, которой обладают люди и ряд животных. Принято считать, что система бинаурального слуха состоит из двух сенсоров (ушей) и системы обработки информации (головной мозг). Данная конфигурация системы позволяет человеку и животным ориентироваться в окружающем мире, общаться между собой, развиваться за счет обучения и самообучения.

**Целью работы** является разработать автономную систему бинаурального слуха и реализовать ее в виде отдельного устройства для применения в робототехнических системах и других областях искусственного интеллекта.

Более конкретно рассмотрим робототехническую систему (РТС), оснащенную сенсорами различной физической природы и смонтированную на подвижной платформе, позволяющей проникать в труднодоступные места [1]. Одной из задач РТС является проведение разведки в местах природных и техногенных катастроф с целью обнаружения живых людей, подающих какие-либо признаки жизни. В частности для этой цели используются микрофоны звукового диапазона (300 – 5000 Гц). Работая с сигналами звукового диапазона, можно решить следующие задачи: определить число источников, определить направления на источники (локация источников), определить природу источника (живой или неживой природы). Имеется множество моделей распространения звука и множество практических реализаций систем локации звука.

В частности, известная трехмерная модель распространения звуковой волны в однородной воздушной среде, находящейся в состоянии покоя и не имеющей резонаторов и ревербераторов, для одного неподвижного источника, создающего звуковое давление $p$, и скоростей распространения волны $u$, $v$ и $w$ по координатам $x$, $y$, $z$ с плотностью воздуха $\rho$, и эластичность воздуха $k$ имеет вид:

$$\rho \frac{du}{dt} + \frac{dp}{dx} = 0; \qquad (1)$$

$$\rho \frac{dv}{dt} + \frac{dp}{dy} = 0; \quad (2)$$

$$\rho \frac{dw}{dt} + \frac{dp}{dz} = 0; \quad (3)$$

$$\frac{dp}{dt} + k\left(\frac{du}{dx} + \frac{dv}{dy} + \frac{dw}{dz}\right) = 0\,. \quad (4)$$

При этом явно предполагается, что источник и приемник(и) имеют вид точки.

Известна [2] и модель пеленгации источника с помощью двух сенсоров, которую геометрически можно интерпретировать схемой как на рис. 1.

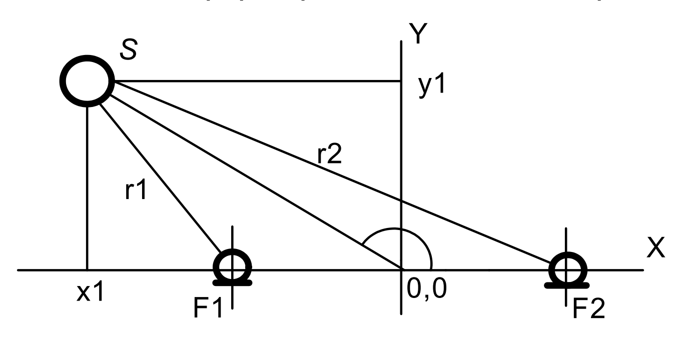



Рисунок 1 – Геометрическая схема модели пеленгации источника

Решение задачи пеленгации сводится к определению расстояний между источником и сенсорами (микрофонами), что эквивалентно измерению разности фаз сигналов на выходе сенсоров для каждого из источников. Общая задача разделяется на ряд подзадач, а именно – разделение сигналов по частотным полосам, определение принадлежности сигналов источникам, фазовые измерения [3], [4].

Для задач обработки сигналов звукового традиционно применяются прямое и обратное преобразования Фурье, которые требуют больших вычислительных ресурсов. Автором предложен [5] другой метод представления сигналов, который значительно понижает требования к вычислительным ресурсам и позволяет решить задачи обработки на обычных микропроцессорах. На основе имеющихся знаний и строится многопроцессорная система реализации бинаурального слуха с массивом микрофонов.

# 1. Пеленгация источника

Обращаясь к схеме на рис. 1, можно видеть, что если расстояние $x_{12}$ между сенсорами *F1* и *F2* фиксировано, то для всех частот звукового диапазона *f* можно вычислить разность фаз:

$$\Delta\varphi = (\varphi 1 - \varphi 2) = F_3\{ x_{12}, f, ( x_1, y_1)\}. \quad (5)$$

Разумеется, пару координат $(x_1, y_1)$ можно представить как функцию разницы расстояний от источника *S* до сенсоров *F1*, *F2* или как угол между источником и осью – *X*, *X*. Из этой модели сразу видна неоднозначность решений – при одном и том же значении частоты имеется минимум 4 варианта правильных решений, каждый из которых лежит в одной из четвертей декартовых координат. Более того, если частота сигнала такова, что ее полупериод несколько раз укладывается в промежуток $x_{12}$, то число решений следует умножить на число поворотов фазы сигнала в промежутке $x_{12}$.

Из этой модели следует, что для правильного построения образа звука для каждого из источников необходимо иметь минимум четыре сенсора давления, разнесенных в пространстве. Из четырех сенсоров $F_0$, $F_x$, $F_y$, $F_z$ образуют пары – $(F_0 + F_x)$, $(F_0 + F_y)$, $(F_0 + F_z)$ (рис. 2).

Каждую пару сенсоров можно рассматривать как стереопару и построить три потока полупериодных образов сигнала относительно центрального сенсора. Далее, используя операции над полупериодами и анализируя фазовые различия сигналов в каждом из трех образовавшихся потоков данных, можно восстановить неискаженный образ исходного сигнала по каждому из заданных направлений или по каждой из устойчивых фаз.

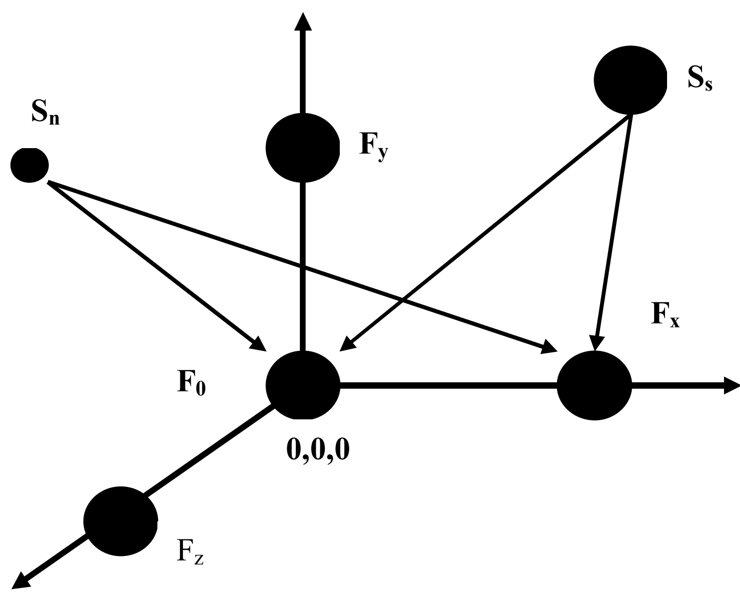



Рисунок 2 – Модель размещения сенсоров в объеме

Из этой модели также следует, что для восстановления образа звука в одной плоскости достаточно трех сенсоров, но достоверность образа будет снижена.

Имеется частный случай, позволяющий восстановить образ звука с помощью двух сенсоров – когда источник звука лежит точно на биссектрисе угла в одной из координатных плоскостей – например, на биссектрисе угла *Y*, *0*, *X* – рис. 2. При этом расстояние от источника *Ss* до сенсоров *Fx*, *Fy* является одинаковым и, следовательно, фазовая задержка сигналов тоже является одинаковой.

В общем случае работа предлагаемого метода для пары сенсоров ($F_0$, $F_x$) состоит в том, что сигнал с сенсора $F_0$ представлен последовательностью полупериодов $P_0$, а сигнал с сенсора *Fx* представлен последовательностью полупериодов $P_x$. В свою очередь, источник звука $S_s$ генерирует последовательность полупериодов $P_x$, а источник шума $S_n$ генерирует последовательность полупериодов $P_n$. В воздушной среде происходит сложение последовательностей полупериодов такое, что $P_0 = P_s + P_n$, $P_x =$ $= P_s + P_n$ *и* $P_0 \neq P_x$.

Задача состоит в том, чтобы по имеющимся $P_0$ и $P_x$ восстановить $P_s$ и $P_n$ и определить направления на источники $S_s$, $S_n$. Задача представления сигналов в виде полупериодов решается применением только целочисленной арифметики, без применения Фурье – анализа. Собственно данные о сигналах представлены в виде целых чисел и для операций над сигналами достаточно целочисленной арифметики, которая встроена на аппаратном уровне в большинство современных микропроцессоров.

Таким образом, решение задачи пеленгации источников сводится к решению задач разделения сигналов по полосам для каждого из сенсоров, вычислению фазового сдвига по полосам сигналов для всех возможных пар сенсоров (6 пар для данного случая).

На рис. 3 показана типовая система обработки звукового сигнала с пары сенсоров. Символами FFT обозначен блок прямого преобразования Фурье, IFFT – обратного преобразования Фурье, IPD – разделение сигналов по фазам. Для *K* пар сенсоров необходимо иметь *K* механизмов сравнения. Соответственно каждый механизм сравнения потребует отдельного микропроцессора, имеющего встроенные преобразования Фурье. В принципе такие микропроцессоры есть (например фирмы Texas Instruments, США), но построение системы, содержащей 6 подобных микропроцессоров, представляется весьма проблематичной задачей. Даже будучи построенной, такая система никак не будет удовлетворять требованиям, предъявляемым к портативным системам по габаритам, энергопотреблению и стоимости.

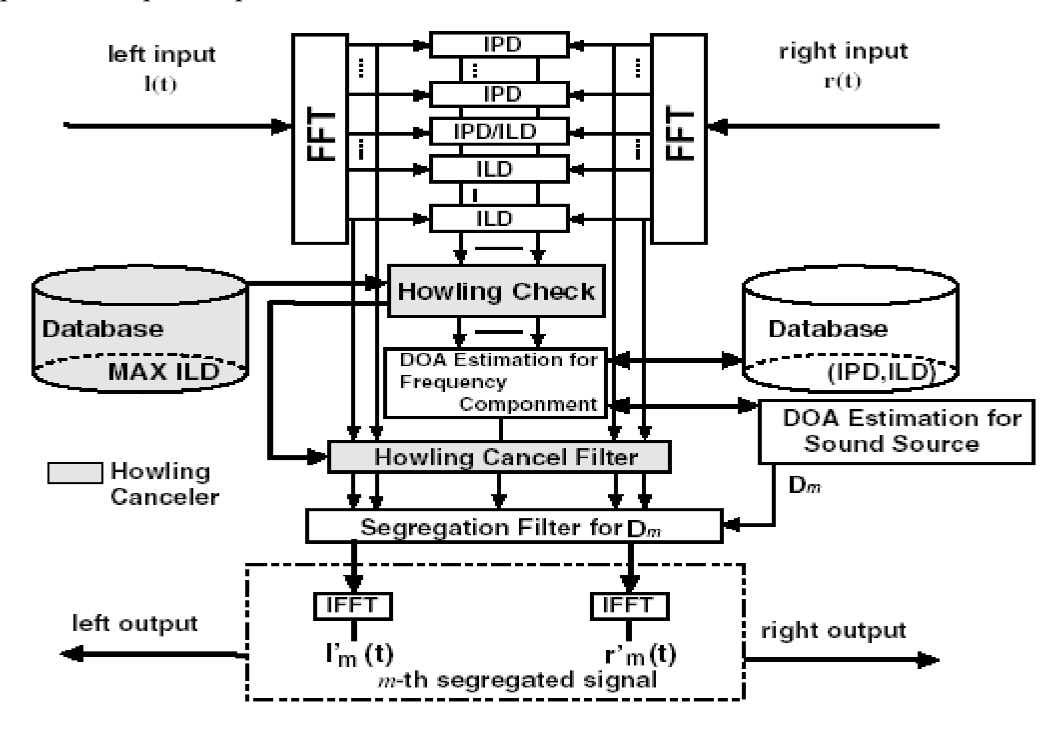



Рисунок 3 – Типовая система обработки сигналов с двух сенсоров

Учитывая специфику работы малогабаритного робота, энергетические и стоимостные показатели, можно сделать вывод о необходимости разработки системы бинаурального слуха, основанную на более простых принципах работы и более простых микропроцессорах.

# 2. Структура системы

При построении системы следует учитывать следующие требования:
– система должна воспринимать звуковые сигналы с расстояния 0,5..5 м с уровнем, достаточным для работы АЦП;
– система должна обладать минимальной инерцией при восприятии сигналов.

Реализация первого требования выполняется применением усилителей с малым уровнем собственных шумов. Реализация второго требования необходима для восприятия слабых сигналов, следующих после сильных сигналов. Известно, что человеческий слух требует некоторого времени на адаптацию для восприятия слабого сигнала после гораздо более мощного сигнала. Применение системы усилителей с автоматической регулировкой усиления (АРУ) в общем решает задачу, но порождает новую проблему –

выбора времени реакции системы. Кроме того, начало сильного сигнала, следующего после слабого, будет неправильно восприниматься до тех пор, пока система АРУ не уменьшит коэффициент усиления. Это грозит потерей сигнала вообще, так как короткий сигнал за время настройки АРУ может просто прекратиться. Например, если рядом с микрофоном с некоторым интервалом падают капли воды (мощный сигнал), то проблематично воспринять слабый сигнал с расстояния 5 м. В качестве системы, не имеющей инерции, можно предложить многоканальную систему, в каждом из каналов устанавливается свой, фиксированный коэффициент усиления и результаты оцифровки каждого канала записываются в отдельный файл данных. На рис. 4 показана структурная схема подобной системы.

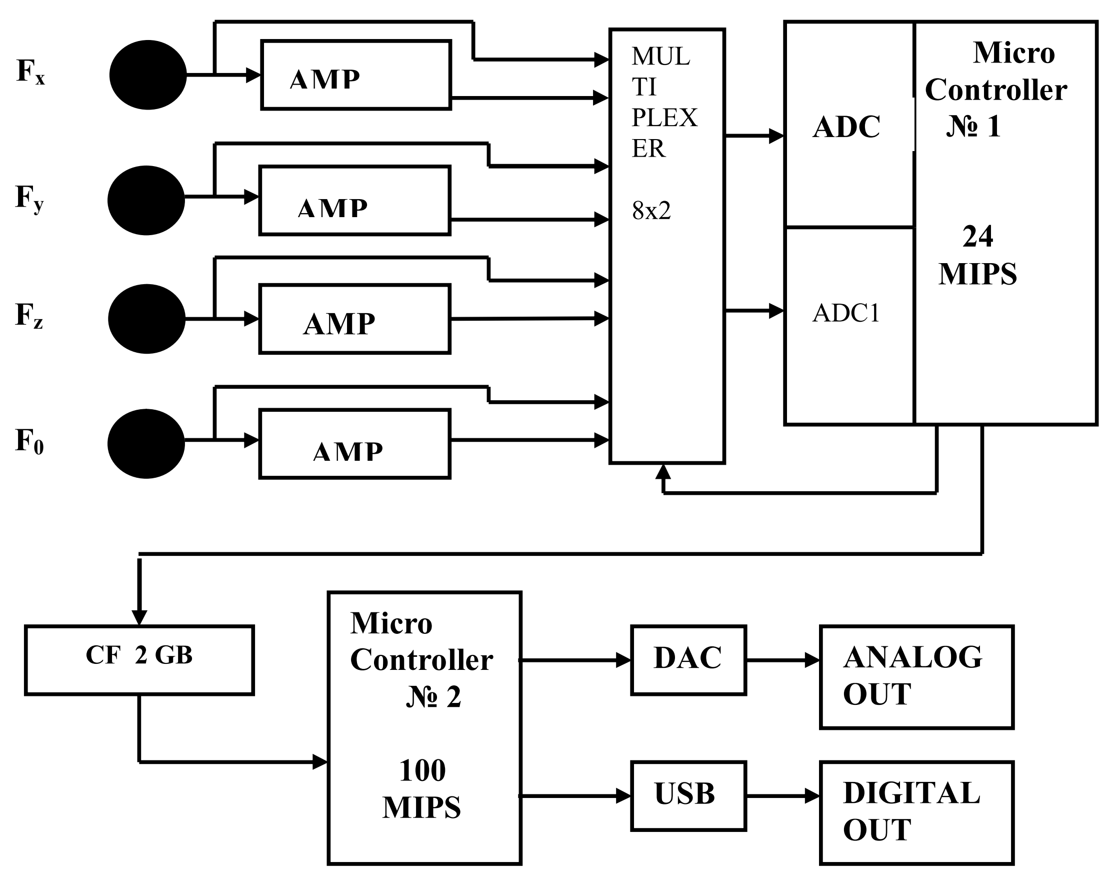



Рисунок 4 – Структурная схема устройства с 4 сенсорами

Собственно микрофоны $F_0$, $F_x$, $F_y$, $F_z$ представляют собой электретные микрофоны от мобильных телефонов, выходы которых подключены ко входам аналогового мультиплексора (коммутатора). К выходам микрофонов подключены входы аналоговых усилителей с постоянным коэффициентом усиления, выходы которых также подключены ко входам аналогового мультиплексора. Выходы аналогового мультиплексора подключены ко входам АЦП микроконтроллера. Собственно два 16-битовых АЦП с частотой дискретизации 1 мГц входят в микроконтроллер № 1 типа C8051F060 (фирма SiLabs, США). Данный микроконтроллер обладает производительностью 24 MIPS (миллионов целых операций в секунду) и достаточным количеством линий ввода – вывода, чтобы управлять переключением входов АЦП и выводить полученные данные в память типа Compact Flash емкостью 2 GB. Данные с АЦП вводятся с фиксированным интервалом равным 5 мкС. Высокая точность АЦП (16 разрядов) и высокая частота дискретизации (до 100 кГц на канал) позволяют сохранить образ сигнала с

минимальными частотными и фазовыми искажениями. Напомним, что входной сигнал имеет диапазон 300 – 5000 Гц, что в 300 – 20 раз менее частоты дискретизации. Можно говорить, что частотные и фазовые искажения, вносимые каналами устройства, не превышают 5 % для любого сигнала в диапазоне 300 – 5000 Гц. Быстродействия микроконтроллера и памяти достаточно, чтобы получить и сохранить данные без потерь, но недостаточно для обработки полученных данных.

Для обработки данных используется микроконтроллер № 2 типа C8051F120 (фирма SiLabs, США). Данный микроконтроллер обладает производительностью 100 MIPS. В задачи этого микроконтроллера входит чтение полученной информации из памяти, обработка этой информации, сохранения результатов обработки в этой же памяти и выдача информации по внешним запросам. Задание режима работы МК № 2 и устройства в целом может производиться как через USB интерфейс, так и с местного пульта (не показан ввиду его тривиальности).

USB интерфейс реализован на микроконтроллере № 3 типа C8051F342 (фирма SiLabs, США). Данный микроконтроллер обладает производительностью 48 MIPS. В задачи этого микроконтроллера входит чтение полученной информации из буферной памяти МК № 2, передача ее в HOST машину, получение управляющих команд из HOST машины и дальнейшая передача управляющей информации МК № 1 и МК № 2.

## Выводы

Предлагаемая многопроцессорная система реализации бинаурального слуха с массивом микрофонов позволяет принципиально реализовать все свойства, присущие слуху человека, – восприятие звуков в широком диапазоне частот и амплитуд, локализацию источников звуков по направлениям, определение подвижности источников. Разумеется, эти свойства реализуются только в ближнем окружении – на расстоянии не более 5 м.

Предлагаемая система обладает рядом свойств, которые не присущи слуху человека:

– благодаря записи всех сигналов в память, система может анализировать и короткие сигналы, например, типа выстрел;
– система имеет время адаптации к сигналу = 0, что позволяет воспринимать без искажений как слабые после сильных, так и сильные после слабых сигналов;
– после разделения сигналов по направлениям и источникам, система может синтезировать сигнал как суммарный от двух и более выбранных источников (например, для выделения общей шумовой картины);
– система может быть быстро обучена поиску определенных сигналов благодаря наличию USB интерфейса. В систему могут быть очень быстро переданы образы сигналов, которые следует обнаруживать или наоборот, игнорировать – то есть обучение системы не требует звукового воспроизведения фильтруемых сигналов.

Применение предлагаемой системы на борту малогабаритного робота позволяет реализовать задачи поиска и локализации источников звуковых сигналов в любых ситуациях и окружающих условиях. Применение предлагаемой системы в автономном варианте позволяет решить вопрос восприятия и разделения по источникам множества голосов, например, во время совещания, дискуссии. Предлагаемая система может служить для распознавания голоса при диалоге двух людей – с ее помощью решается задача получения качественного сигнала без применения микротелефонных гарнитур и в широком диапазоне расстояний между говорящими. Предлагаемая система может

служить для распознавания голоса в системах голосового управления оборудованием в промышленности (в т.ч. управления роботами), быту – управляющий голос воспринимается со значительного расстояния без применения носимого микрофона. Предлагаемая система может служить как узконаправленный микрофон, то есть воспринимать сигнал только с заданного направления.

# Литература

1. Поливцев С.А., Шевченко А.И. Патент Украины № 75991. Шагающий механизм малогабаритного робота. Выдан 15.06.2006.
2. Понтрягин Л.С. Метод координат. – Москва, 2004. – 135 с.
3. Azimuthal and elevation localization of two sound sources using interaural phase and level differences. Yoshifumi Chisaki1;_, Sho Kawano1, Kyoko Nagata1, Kotaro Matsuo1, Hidetoshi Nakashima2;y and Tsuyoshi Usagawa1; Acoust. Sci. & Tech. 29, 2 (2008).
4. POST FILTER FOR MICROPHONE ARRAY. Inventor: AKAGI MASATO (JP); LI JUNFENG (JP); (+2) Applicant: JAPAN ADVANCED INST OF SCIENCE (JP); TOYOTA MOTOR CO LTD (JP). Publication info: EP1931169-2008-06-11.
5. Поливцев С.А., Кобыляков В.В. Метод полупериодного анализа для массива микрофонов // Искусственный интеллект. – 2006. – № 4. – С. 787-795.

***С.О. Полівцев***

**Багатопроцесорна система реалізації бінаурального слуху із масивом мікрофонів**

У статті пропонується варіант практичної реалізації системи бінаурального слуху, властивого людині та багатьом представникам тваринного світу. Властивості бінаурального слуху реалізуються шляхом операцій над образами акустичних сигналів, які надходять із масиву мікрофонів. Реалізація передбачає зв'язну, синхронізовану роботу 3 мікропроцесорів, що виконують читання сигналів з 4 мікрофонів через два 16-розрядних АЦП із частотою дискретизації 100кГц по кожному каналу із синхронізацією між каналами з точністю 5 мкС. У реалізацію включена Flash карта місткістю 2 GB та мікропроцесор, що реалізує інтерфейс USB із персональної ЕОМ. Реалізація передбачає як автономну роботу системи, так і сумісно з персональною ЕОМ. Система має властивості, які не властиві слуху людини, – нульовий час адаптації при переході від сильного сигналу до слабкого та навпаки, локалізацію дуже коротких сигналів.

***S.A. Polivtsev***

**Multiprocessors System of Realization Binaural Hearing with a Array of Microphones**

In article the variant of practical realization of system binaural hearing inherent is offered to the person and many representatives of fauna. Properties binaural hearing are realized by operations above images of the acoustic signals received from a array of microphones. Realization assumes the coherent, synchronized work of 3 microprocessors which are carrying out reading of signals from 4 microphones through two 16 digit ADC with frequency of digitization 100 kHz on each channel with synchronization between channels with accuracy 5 uS. In realized the Flash card in capacity 2 GB and the microprocessor realizing interface USB from the personal COMPUTER is included. Realization provides independent work of system and together with the personal COMPUTER. The system has properties not inherent in hearing of the person – zero time of adaptation at transition from a strong signal to weak and on the contrary, localization of very short signals.

*Статья поступила в редакцию 08.12.2008.*